

ANALISIS PENGELOMPOKAN PENYEBARAN LULUSAN MAHASISWA UNIVERSITAS GUNADARMA MENGGUNAKAN METODE CLUSTERING

Sheila Pramita Hervianti

Program Studi Magister Sistem Informasi Universitas Gunadarma
Jln. Kenari III/5 Jakarta Pusat, 087788116730

E-mail : sheilapramitah@gmail.com

ABSTRAK

Teknologi database saat ini memungkinkan untuk menyimpan sejumlah data dalam jumlah yang sangat besar dan terakumulasi namun disinilah awal timbulnya persoalan ledakan data, contohnya pada Universitas. Universitas Gunadarma masih melakukan analisis data secara manual untuk menentukan rencana strategi promosi dari sebaran mahasiswa berdasarkan geografis. Sebaran mahasiswa dapat dilakukan dengan cara mengelompokkan data mahasiswa berdasarkan kesamaan ciri dari data menggunakan metode *clustering* dengan algoritma k-mean. Pembuatan aplikasi *data mining* bertujuan untuk memudahkan menganalisis pengelompokan sebaran mahasiswa. Data yang digunakan adalah data angkatan mahasiswa tahun 2008 sampai tahun 2010 yang telah lulus. Pada penelitian ini aplikasi dibangun dengan menggunakan perangkat lunak java dan MySQL. Atribut yang digunakan adalah provinsi asal, program studi IPK dan lama studi mahasiswa. *Cluster* mahasiswa yang terbentuk adalah tiga *cluster*, dengan *cluster* pertama 6.686 mahasiswa, *cluster* kedua 3.938 mahasiswa dan *cluster* ketiga sejumlah 3.258 mahasiswa. Didapat pula hasil clustering setiap provinsi berdasarkan IPK dan lama studi mahasiswa. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan *cluster* yang terbentuk oleh pihak Universitas Gunadarma.

Kata Kunci: Analisis, Clustering, *K-means*, Mahasiswa, WEKA, Promosi.

1. PENDAHULUAN

Latar Belakang

Selama dua decade terakhir telah terjadi peningkatan yang dramatis terhadap jumlah informasi atau data yang disimpan secara elektronik. Teknologi database saat ini memungkinkan untuk menyimpan sejumlah data dalam jumlah yang sangat besar dan terakumulasi namun disinilah awal timbulnya persoalan ledakan data, contohnya pada Universitas.

Universitas Gunadarma merupakan salah satu universitas swasta yang terkenal di Indonesia sehingga mahasiswanya banyak yang berasal dari berbagai daerah di Indonesia dengan begitu pihak universitas memproses ribuan data mahasiswa setiap tahunnya. Pihak universitas menggunakan sistem manajemen basis data tradisional dalam pengelolaan datanya (Didin Mukodim, 2015). Data selama 3 tahun terakhir dimanfaatkan oleh pihak institusi untuk mengetahui sebaran mahasiswa Universitas Gunadarma berdasarkan provinsi dalam kegiatan promosi (Didin Mukodim, 2015). Sistem manajemen basis data tradisional sangat baik dalam menyimpan data secara cepat, aman dan efisien ke dalam basis data namun tidak cukup baik dalam hal kemampuan melakukan analisa terhadap data-data yang ada

sehingga Universitas Gunadarma masih melakukan analisis data secara manual untuk menentukan rencana strategi promosi dari sebaran mahasiswa berdasarkan geografis (Didin Mukodim, 2015). Disinilah peran *data mining* akan memberikan kontribusi dalam membantu pengolahan data dan analisis data.

Penggunaan aplikasi atau perangkat lunak *data mining* dapat membantu pihak Universitas Gunadarma dalam melakukan analisis data dan menentukan rencana strategi promosi dari sebaran mahasiswa berdasarkan geografis. Sebaran mahasiswa dapat dilakukan dengan cara mengelompokkan data mahasiswa berdasarkan kesamaan ciri dari data. Salah satu metode yang terkenal pada *data mining* untuk mengelompokkan data adalah dengan metode analisis *cluster* atau *clustering*.

Metode analisis *cluster* atau *clustering* baik diterapkan sebagai model analisis data otomatis pada aplikasi. Analisis cluster mampu melakukan pengorganisasian kumpulan pola ke dalam kelompok-kelompok (*cluster*) berdasarkan kesamaan ciri atau sifat dari pola-pola. Analisis cluster menemukan kelompok mahasiswa dengan perilaku atau ciri serupa pada *database* mahasiswa yang

mengandung sifat-sifat dan informasi pada masa lalu. Selanjutnya, rencana strategi promosi dapat ditentukan melalui segmentasi yang didukung proses pengelompokan untuk menemukan batas antara kelompok-kelompok mahasiswa.

Berdasarkan latar belakang, penulis tertarik untuk mengangkat topik penelitian berjudul “**Analisis Penyebaran Pengelompokan Lulusan Mahasiswa Universitas Gunadarma dengan Metode Clustering**”. Hasil dari analisis penyebaran pengelompokan lulusan mahasiswa Universitas Gunadarma didapat dari pembuatan dan implementasi aplikasi *data mining* yang menerapkan metode analisis *cluster* atau *clustering* dengan menggunakan algoritma K-means. Diharapkan dengan adanya aplikasi ini dapat membantu pihak institusi khususnya bagian sekretariat Universitas Gunadarma dalam melakukan analisis data dan menentukan rencana strategi promosi dari sebaran mahasiswa.

Tujuan Penelitian

Adapun yang menjadi tujuan dari penelitian adalah membangun aplikasi *data mining* dengan menerapkan metode analisis *cluster* atau *clustering* menggunakan algoritma K-means dan menganalisis hasil aplikasi berupa pengelompokan sebaran mahasiswa Universitas Gunadarma.

2. METODELOGI

Pengumpulan Data

Pada tahap pengumpulan data, penulis mengumpulkan informasi atau data yang dibutuhkan agar dapat diolah dan dianalisis. Pengumpulan data yang dilakukan penulis dengan menggunakan metode dokumen dan wawancara.

Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak yang digunakan dalam membangun aplikasi *data mining* menggunakan metode *System Development Life Cycle*.

1. Perencanaan Aplikasi

Pembuatan aplikasi *data mining* yang menerapkan metode analisis *cluster* atau *clustering* menggunakan algoritma K-means bertujuan untuk menganalisis pengelompokan sebaran mahasiswa Universitas Gunadarma.

2. Analisis Sistem

Tahapan analisis terdiri dari tiga bagian yaitu analisis kebutuhan *input*, *output* dan analisis proses *clustering*.

a. Analisis Kebutuhan Input

Input atau masukan dari aplikasi *data mining* yang menerapkan metode analisis *cluster* atau *clustering* dengan menggunakan algoritma K-means adalah input data mahasiswa dan input nilai K.

b. Analisis Kebutuhan Output

Hasil keluaran (*output*) dari aplikasi *data mining* yang menerapkan metode analisis *cluster* atau *clustering* menggunakan algoritma K-means berupa informasi mengenai kelompok - kelompok mahasiswa yang telah dikelompokkan menggunakan algoritma *k-means* berdasarkan jurusan, IPK, lama kuliah dan provinsi.

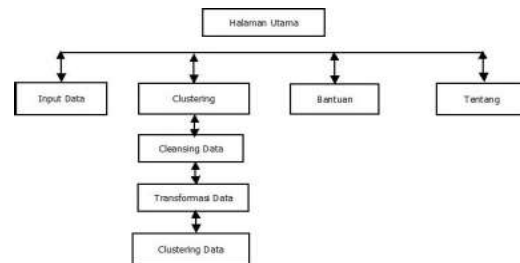
c. Analisis Proses Clustering

Standarisasi proses *data mining* yang digunakan pada penelitian ini adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*).

3. Perancangan Sistem Aplikasi

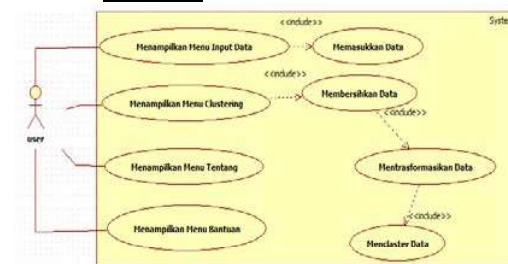
Tahap perancangan aplikasi akan dimulai dengan membahas mengenai perancangan sistem aplikasi menggunakan struktur navigasi, *Unified Modelling Language* (UML), rancangan database dan rancangan tampilan.

a. Perancangan Struktur Navigasi

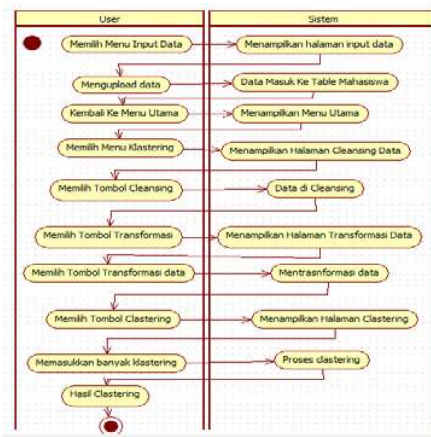


Gambar 3.5 Struktur Navigasi Hirarki

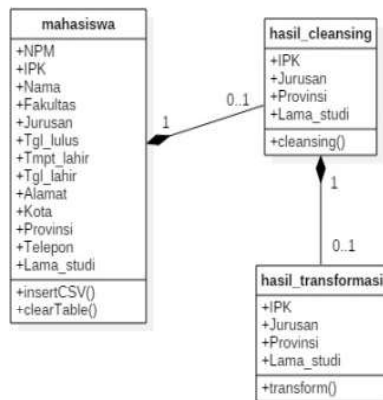
b. Perancangan Unified Modeling Language



Gambar 3.6 Use Case Diagram



Gambar 3.7 Activity Diagram



Gambar 3.8 Class Diagram

Rancangan Interface

Pada tahap ini penulis akan menguraikan penjelasan rancangan *interface* yang terdapat pada aplikasi. Perancangan *interface* dimaksudkan untuk pembuatan rancangan tampilan. Rancangan *interface* aplikasi bertujuan untuk membantu penulis dalam membuat aplikasi karena sebelum membuat aplikasi dibutuhkan sebuah rancangan untuk memahami struktur aplikasi.

Rancangan Database

Perancangan *database* adalah proses untuk menentukan isi dan pengaturan data yang dibutuhkan untuk mendukung berbagai rancangan sistem. Pada aplikasi ini terdiri dari 3 tabel data yaitu tabel mahasiswa, tabel hasil_cleansing, tabel hasil_transformasi.

3. LANDASAN TEORI

Data Mining

Data mining adalah ekstraksi pola yang menarik dari data dalam jumlah besar (Han, 2000). Suatu pola dikatakan menarik apabila pola tersebut

tidak sepele, implisit, tidak diketahui sebelumnya, dan berguna (Han, 2000).

Pengklasteran (Clustering)

Teknik atau metode pengklasteran termasuk teknik yang sudah cukup dikenal dan banyak dipakai sebagai alat atau mesin pembelajaran (Prasetyo, 2014). Usaha-usaha untuk menghitung jumlah kluster yang optimal dan pengklasteran yang paling baik masih terus dilakukan.

Algoritma K-Means

Algoritma *k-means* merupakan algoritma yang membutuhkan parameter input sebanyak k dan membagi sekumpulan n objek kedalam k cluster sehingga tingkat kemiripan antar anggota dalam satu cluster tinggi sedangkan tingkat kemiripan dengan anggota pada cluster lain sangat rendah. Kemiripan anggota terhadap cluster diukur dengan kedekatan objek terhadap nilai *mean* pada cluster atau dapat disebut sebagai *centroid cluster*.

Tahapan Data Mining

Dalam melakukan *data mining*, terdapat sebuah panduan yang banyak dipakai dalam dunia industri, yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM). CRISP-DM dikembangkan tahun 1996 oleh analis dari beberapa industri seperti Daimler Chrysler, SPSS dan NCR. CRISP-DM menyediakan standar proses *data mining* sebagai strategi pemecahan masalah secara umum dari bisnis atau unit penelitian. Dalam CRISP-DM, sebuah proyek *data mining* memiliki siklus hidup yang terbagi dalam enam fase.

UML

Unified Modelling Language (UML) menawarkan sebuah standar untuk merancang model sebuah sistem. Tujuan UML adalah :

1. Memberikan model yang siap pakai, bahasa pemodelan visual yang ekspresif untuk mengembangkan dan saling menukar model dengan mudah dan dimengerti secara umum.
2. Memberikan bahasa pemodelan yang bebas dari berbagai bahasa pemrograman dan proses rekayasa.
3. Menyatukan praktik-praktik terbaik yang terdapat dalam pemodelan.

UML menyediakan beberapa *notasi dan artifact standar* yang bisa digunakan sebagai alat komunikasi bagi para pelaku dalam proses analisis dan desain. Artifact didalam UML didefinisikan sebagai informasi dalam bentuk yang digunakan atau dihasilkan dalam proses pengembangan perangkat. Contohnya adalah source code yang dihasilkan oleh proses pemrograman.

Flowchart

Flowchart atau diagram alir merupakan sebuah diagram dengan simbol garis yang menyatakan aliran algoritma atau proses yang menampilkan langkah-langkah yang disimbolkan dalam bentuk kotak, beserta urutannya dengan menghubungkan masing-masing langkah tersebut menggunakan tanda panah. Diagram ini bisa member solusi selangkah demi selangkah untuk penyelesaian masalah yang ada dalam proses atau algoritma tersebut. Bagan alir (*flowchart*) adalah bagan (*chart*) yang menunjukkan alir (*flow*) di dalam program atau prosedur sistem secara logika. Bagan alir program (*program flowchart*) merupakan bagan alir yang mirip dengan bagan alir sistem, yaitu untuk menggambarkan prosedur di dalam sistem. (Jogiyanto, 2005)

Struktur Navigasi

Navigasi (penjejakan) merupakan struktur dari suatu program dan paling penting dalam pembuatan situs. Peta penjejakan mempunyai beberapa ciri khas yang dapat digolongkan menurut kebutuhan objek, kemudahan pemakaian, keinteraksiannya dan kemudahan pembuatannya.

Weka

WEKA merupakan API Java yang menyediakan API untuk pengolahan dalam *data mining* yang berbasis *open source* (GPL) dan berengine JAVA. WEKA dikembangkan oleh Universitas Waikato di Selandia Baru dan juga merupakan perangkat lunak gratis yang tersedia dibawah GNU (*General Public License*). WEKA menyediakan penggunaan teknik klasifikasi menggunakan pohon keputusan dengan algoritma J48. Teknik yang digunakan WEKA adalah *classifier* (Han, 2000).

4. HASIL DAN PEMBAHASAN

Pada tahap ini penulis melakukan proses pembangunan sistem secara keseluruhan berdasarkan hasil rancangan yang telah dibuat dan selanjutnya akan diimplementasikan dalam bentuk pemrograman komputer. Tahap selanjutnya merupakan proses pengujian sistem yang akan dilakukan menggunakan metode *black box testing*.

Implementasi

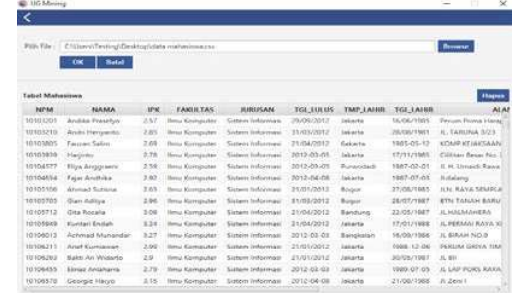
Tahapan awal dalam pembuatan aplikasi Berikut adalah tahapan pembuatan UG mining. Pada tahap awal implementasi aplikasi diperlukan instalasi *software* untuk membuat aplikasi yang dapat berjalan pada sistem operasi *Windows*. *Software* yang digunakan adalah NetBeans IDE dan XAMPP untuk sistem operasi *Windows*.

a. Pembuatan Halaman Utama



Gambar 4.2 Tampilan Halaman Utama

b. Pembuatan Halaman Upload File



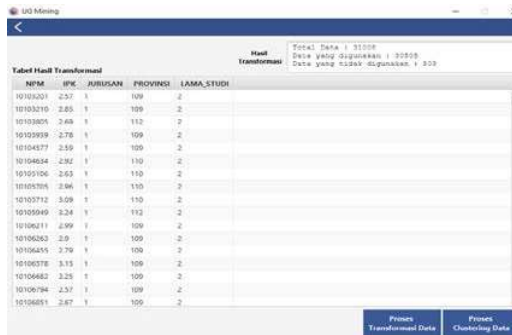
Gambar 4.3 Tampilan Halaman Upload File

c. Pembuatan Halaman Cleansing



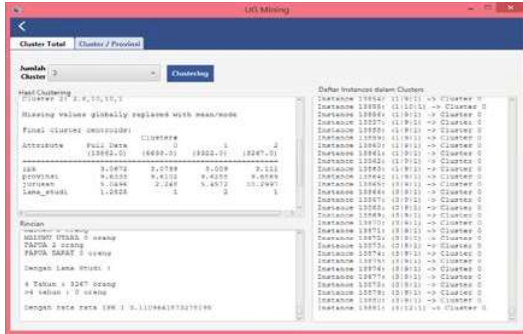
Gambar 4.4 Tampilan Halaman Cleansing

d. Pembuatan Halaman Transformasi



Gambar 4.5 Tampilan Halaman Transformasi

e. Pembuatan Halaman Clustering



Gambar 4.6 Tampilan Halaman Clustering



Gambar 4.6 Tampilan Halaman Tabel Hasil Clustering

f. Pembuatan Halaman Bantuan



Gambar 4.7 Tampilan Halaman Bantuan

g. Pembuatan Halaman Tentang



Gambar 4.8 Tampilan Halaman Tentang

Pengujian Perhitungan Analisis Proses Clustering

Hasil perhitungan telah didapatkan dari sistem, penulis akan melakukan uji coba dengan perhitungan yang dilakukan secara manual. Pada tabel 4.6 merupakan sebagian dari data mahasiswa yang ada pada database akademik pada aplikasi UG Mining.

1. Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan mejadi tiga *cluster*.
2. Tentukan titik pusat awal dari setiap *cluster*. Dalam penelitian ini titik pusat awal ditentukan secara *random* dan didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel 4.7.
3. Langkah selanjutnya adalah menempatkan setiap data ke *cluster* dengan menghitung jarak terdekat antar objek dan *centroid* menggunakan persamaan *Euclidian Distance* seperti berikut :

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2 + (p_4 - q_4)^2}$$

Pertama akan dihitung jarak dari data mahasiswa pertama ke pusat *cluster* pertama.

$$d(1,1) = \sqrt{(112 - 109)^2 + (1 - 1)^2 + (2,69 - 2,57)^2 + (2 - 2)^2}$$

$$= \sqrt{(3)^2 + (0)^2 + (0,12)^2 + (0)^2}$$

$$= \sqrt{9 + 0 + 0,0144 + 0}$$

$$d(1,1) = 3.0023$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 3.0023. Jarak data mahasiswa pertama ke *cluster* kedua :

$$d(1,2) = \sqrt{(112 - 110)^2 + (1 - 8)^2 + (2,69 - 3,18)^2 + (2 - 1)^2}$$

$$= \sqrt{(2)^2 + (-7)^2 + (-0,49)^2 + (1)^2}$$

$$= \sqrt{4 + 49 + 0,24 + 1}$$

$$d(1,2) = 7,3647$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* kedua adalah 7.3647. Jarak data mahasiswa pertama ke *cluster* ketiga :

$$d(1,3) = \sqrt{(112 - 109)^2 + (1 - 5)^2 + (2,69 - 3,58)^2 + (2 - 1)^2}$$

$$= \sqrt{(3)^2 + (-4)^2 + (-0,89)^2 + (1)^2}$$

$$= \sqrt{9 + 16 + 0,7921 + 1}$$

$$d(1,3) = 10,8797$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* ketiga adalah 10.8797. Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data mahasiswa pertama yang paling dekat adalah *cluster* 1, sehingga mahasiswa pertama dimasukkan ke dalam *cluster* 1.

Data ke-	Provinsi	Jurusan	IPK	Lama Studi	C1	C2	C3
1	112	1	2.69	2	3.0023	7.3647	5.1761
2	110	11	2.55	2	10.0498	3.2244	6.2498
3	109	3	3.14	1	2.3075	5.0991	2.0478
4	109	4	3.27	1	3.2388	4.1240	1.0469
5	109	10	3.46	1	9.0990	2.2535	5.0014
6	106	10	3.70	1	9.6060	4.5022	5.8321
7	118	10	3.32	1	12.7891	8.2473	10.2989

Tabel 4.8 Hasil Perhitungan Setiap *Cluster*

5. KESIMPULAN

Aplikasi pengelompokan persebaran mahasiswa Universitas Gunadarma dengan menggunakan *clustering* k-means telah selesai dibuat. Pemrograman aplikasi pengelompokan persebaran mahasiswa Universitas Gunadarma dengan menggunakan *clustering* k-means ini dibuat menggunakan bahasa pemrograman Java dan memanfaatkan library yang dimiliki WEKA untuk proses *clustering*. Aplikasi ini digunakan untuk mengetahui pengelompokan persebaran mahasiswa universitas gunadarma agar dapat membantu dalam kegiatan promosi. Aplikasi ini telah di uji coba menggunakan metode black box.

Hasil pengujian menggunakan metode black box untuk fungsi halaman, semua halaman sukses ditampilkan sesuai harapan. Hasil pengujian *clustering* setiap provinsi berhasil di tampilan dan dapat dibaca dengan mudah. Berdasarkan hasil uji coba dan analisa untuk setiap provinsi, persebaran mahasiswa terbanyak berada pada provinsi DKI Jakarta dengan kelulusan sesuai target yaitu 4 tahun dan IPK di atas 3.1. Provinsi Yogyakarta, Nusa Tenggara Barat dan Sulawesi Tengah memiliki persebaran yang sangat minimum dan untuk provinsi Gorontalo, Sulawesi Barat, Maluku Utara dan Papua Barat belum ada persebaran mahasiswa Universitas Gunadarma.

DAFTAR PUSTAKA

- Agusta, Y. (2007). K-means – Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*, Vol. 3.
- Arlinda, Zefy. (2012). *Implementasi Algoritma K-means Untuk Mengetahui Sebaran Mahasiswa Sebagai Informasi Penunjang Promosi Universitas*. Skripsi Fakultas Teknik Informatika: Universitas Bengkulu.
- A.S, Rosa, dan Shalahuddin,M. (2011). *Modul Pembelajaran Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)*. Bandung: Penerbit Modula Bandung.
- Dias, Rima. (2014). *Data Mining Menggunakan Algoritma K-means Clustering Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro*. Skripsi Fakultas Ilmu Komputer : Universitas Dian Nuswantoro.
- Han, J., Pei, J., and Yin, Y. (2000). *Mining frequent patterns without candidate generation*. In *SIGMOD '00*.
- Heryanto, Imam. (2012). *Mudah Belajar Java Revisi Kedua*. Bandung : Informatika Bandung.
- Hoffer, Ramesh and Topi. (2012). *Modern Database Management*. Harlow United Kingdom : [Pearson Education Limited](http://www.pearsoned.com).
- Jogiyanto, Hartono. (2005). *Analisis dan Desain Sistem Informasi, Edisi III*. Yogyakarta : Penerbit ANDI.
- Kusrini dan luthfi taufiq Emha. (2009). *Algoritma Data Mining*, Yogyakarta : Penerbit ANDI.
- Oscar, Johan. (2013). Implementasi Algoritma K-Menas Clustering Untuk Menentukan Strategi Marketing President University. *Jurnal Ilmiah Teknik Industri Vol. 12*.
- Prasetyo, Eko. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta : Penerbit ANDI.
- Rizky. (2011). *Konsep Dasar Rekayasa Perangkat Lunak*. Jakarta : PT Prestasi Pustakarya.

- Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu
- Saputra, dan Feni Agustin. (2012). *Membangun Sistem Aplikasi E-commerce dan SMS*. Jakarta : PT Elex Media Komputindo.
- Sulianta, Feri. (2008). *Seri Referensi Praktis Manajemen IT*. Jakarta : PT Elex Media Komputindo.
- Widodo, Prabowo. (2011). *Menggunakan UML*. Bandung: Informatika Bandung